RESEARCH NOTE

Population genetic structure based on SSR markers in alfalfa (*Medicago sativa* L.) from various regions contiguous to the centres of origin of the species

M. FALAHATI-ANBARAN[1,2*], A. A. HABASHI[2*], M. ESFAHANY[3], S. A. MOHAMMADI[2,4] and B. GHAREYAZIE[2]

[1]*Agricultural Biotechnology Research Institute of North Region of Iran (ABRINI), Rasht Ghazvin Road, Rasht 41635-4115, Iran*
[2]*Agricultural Biotechnology Research Institute of Iran (ABRII), Mahdasht Road, Tehran, Karaj 31535-1897, Iran*
[3]*Department of Agronomy and Plant Breeding, Faculty of Agriculture, University of Guilan, Rasht 41635-1314, Iran*
[4]*Department of Crop Production and Breeding, Faculty of Agriculture, University of Tabriz, Tabriz 51666, Iran*

Introduction

Alfalfa (*Medicago sativa* L.) is one of the most important forage crops in the world, and is believed to have originated in the Caucasus region: northeastern Turkey, Turkmenistan and northwestern Iran (Michaud *et al*. 1988). Cultivated alfalfa is autotetraploid ($2n = 4x = 32$) (McCoy and Bingham 1988), cross-pollinated (allogamous) and seed propagated. Severe inbreeding depression and tetrasomic inheritance make it difficult to carry out many types of population-genetic studies in this species. Therefore, breeding approaches to develop new (synthetic) cultivars usually consist of intermating a number of superior selected individual plants in a crossing block. One way to identify the maximally diverse parental genotypes is through an evaluation of genetic diversity using molecular markers. In addition, the analysis of genetic diversity in outbreeding forage species such as alfalfa is an important step for cultivar identification, seed purity analysis, and germplasm management.

Microsatellites, also known as simple sequence repeats (SSR), consist of a 2–5-nucleotide core unit that is tandemly repeated; these repeats are randomly interspersed in eukaryotic genomes (Wang *et al*. 1994) and are codominantly inherited. Ten SSR loci have been identified and developed for genetic mapping in diploid and tetraploid alfalfa (Diwan *et al*. 1997, 2000). No attempts have been made so far to characterize the widely used Iranian alfalfa populations using SSR markers. These populations are routinely differentiated using mosphological descriptors, and although such descriptions are indeed useful from a breeding perspective, they are inadequate for analysis of population genetic structure. Because alfalfa is native to northwestern Iran, a high level of genetic diversity may be expected among individual plants within and between populations. Thus, the objectives of the present study were to (a) identify polymorphic SSR markers with high informative value for the Iranian alfalfa in terms of population differentiation and identification, (b) estimate the genetic population structure, and (c) compare the genetic diversity between different geographical regions, especially across different latitudes.

The average within-population genetic diversity evaluated using eight simple sequence repeat (SSR) markers ranged from 0.82 in the Kerman population (low latitudes) to 0.93 in the Ghareyonjeh population (higher latitudes). The genetic relationships among populations were analysed using cluster analysis by UPGMA algorithm based on the coancestry coefficient matrix. The higher distance between Ghareyonjeh population and others was due to the presence of a specific allele in all plants within this population. The trend of changes in genetic diversity across populations was concordant with latitude changes. The obtained results suggest that SSR markers are useful for identification of alfalfa populations, for assessing interpopulation and intrapopulation genetic diversity, and for estimation of genetic differentiation.

Material and methods

Plant material

Six cultivated alfalfa populations from different geographical regions of Iran, together representing a majority of the

*For correspondence. E-mail: falahati_m@abrii.ac.ir; habashia@yahoo.com.

Keywords. simple sequence repeats; crop genetics; genetic diversity; alfalfa; *Medicago sativa*.

cultivated acreage of alfalfa in the country, were selected for this study (figure 1). The seed material was obtained from the National Plant Gene Bank of Iran, Karaj. Ten plants per population were randomly selected, and green healthy leaves from each plant were collected for DNA extraction.

Figure 1. The geographical distribution of Iranian alfalfa populations, their centres of origin, and level of genetic diversity (GD) in different latitudes. Genetic diversity increases from south to northwestern Iran.

SSR analysis

Total genomic DNA was extracted following Dellaporta *et al*. (1983) with minor modifications. Eight primer pairs detecting SSR loci in alfalfa, as described by Diwan *et al*. (2000), were used in the amplification reactions (see table 1). The PCR reactions were performed in a mixture containing 10 ng template DNA, 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl buffer pH 8.3, 0.1% gelatin, 0.2 mM of each dNTP, 20 pmol of each primer, and 1 U *Taq* DNA polymerase in a total reaction volume of 15 μl using the following PCR profile: 94°C for 5 min; 10 initial cycles as touchdown PCR; 25 subsequent cycles at 94°C for 30 s, 52°C for 30 s, 72°C for 30 s, and a final extension at 72°C for 5 min. The PCR products were mixed with an equal volume of formamide dye (98%), and then 5 μl of the sample was loaded onto standard 6% polyacrylamide denaturing gels, run at 90 W for 30–40 min in 1× TBE buffer, and visualized by silver staining (Bassam *et al*. 1991).

Statistical analysis

Fragments amplified by SSR primers were scored as 1 or 0 for presence or absence of bands, respectively. The average genetic diversity was estimated by computing the mean number of pairwise differences over all loci. Pairwise F_{ST} matrix

Table 1. Population-genetic data for eight SSR loci in six alfalfa populations from Iran.

Locus	Number of alleles		Range of allele size (bp)		Average number of alleles per plant (range)		PIC		F_{ST}	Number of different alleles per plant			
	Previous studies*	This study	Previous studies	This study	Previous study	This study	Subtropical	Semi-cold		1	2	3	4
AFct32	14 (M) 14 (D)	14	132–186 (M) 100–188 (D)	103–166	2.8 (1–4) (M)	2.63 (1–4)	0.81	0.84	0.0017	0.08	0.36	0.41	0.15
AFca11	9 (M) 10 (D)	12	199–195 (M) 133–177 (D)	139–188	2.8 (1–4) (M)	2.57 (1–4)	0.83	0.83	–0.0066	0.07	0.40	0.43	0.10
AFca16	-	8	-	89–103	-	2.7 (2–4)	0.76	0.76	0.0079	0.00	0.40	0.50	0.10
AFctt1	14 (M) 11 (D)	11	100–202 (M) 102–135 (D)	100–130	3.1 (1–4) (M)	2.35 (1–4)	0.81	0.80	0.0098	0.13	0.44	0.38	0.05
AFat15	-	6	-	147–169	-	1.92 (1–4)	0.58	0.62	0.0482	0.25	0.60	0.13	0.02
AFct45	-	7	-	132–144	-	2.00 (1–3)	0.74	0.78	0.0341	0.12	0.76	0.12	0.00
AFct11	-	6	-	186–196	-	1.60 (1–4)	0.51	0.70	0.0914	0.54	0.35	0.07	0.04
MTLEC2A	7 (D)	4	136–236 (D)	180–189	-	2.32 (1–4)	0.66	0.60	0.0183	0.12	0.46	0.40	0.02
Average	-	8.5	-	-	2.9	2.26	0.71	0.74	0.024	0.16	0.47	0.31	0.06

*D, Diwan *et al.* 1997; M, Mengoni *et al.* 2000.
PIC value at AFct11 locus was significantly different between subtropical and semi-cold region ($P < 0.05$).

and hierarchical analysis of molecular variance (AMOVA) (Excoffier *et al*. 1992) were carried out using ARLEQUIN 2.000 software (Schneider *et al*. 2000). The pairwise F_{ST} matrix was used for cluster analysis. Cluster analysis and calculation of genetic similarity were performed using the NTSYS-pc 2.02 package (Rohlf 1998). Allelic polymorphic information content (PIC) was calculated using the formula $PIC = 1 - \Sigma(P_{ij})^2$, where P_{ij} is the frequency of the *i*th allele in the *j*th population, for each SSR locus (Botstein *et al*. 1980).

Results

A total of 68 alleles were detected across all eight SSR primer pairs, with the number of alleles per SSR locus ranging from 4 to 14 (table 1). The majority of the loci scored in this study were highly polymorphic, having 8.5 alleles per locus on average. The PIC value for each locus was also estimated in two population subsets, representing the subtropical and semi-cold regions (table 1). The PIC values ranged from 0.60 at AFct11 and AFat15, to 0.83 at AFca11. PIC values at five SSR loci—AFct32, AFct45, AFct11, AFat15 and MTLEC2A—differed between semi-cold and subtropical regions, but only the difference between regions for AFct11 was statistically significant ($P < 0.05$). *F* statistics over the six populations were also estimated for each SSR locus (table 1). The average number of alleles observed per plant at each locus ranged from 1.60 in AFct11 to 2.7 in AFca16, and the average for all loci was 2.26. The proportion of plants containing one allele or two, three and four different alleles at each locus is given in table 1, and the frequency of plants with either one allele or four different alleles was less than those with two and three alleles in all populations. AFca16 was the most informative locus for population identification, with one specific 89-bp-long allele being present in all genotypes of the Ghareyonjeh population (Hamedan ecotypes), which was not detected in other populations. This specific allele was also observed in Ardabil ecotype of Ghareyonjeh population.

Interpopulation and intrapopulation genetic diversity

A high level of genetic diversity (average 0.89) at all the SSR loci was found in all Iranian alfalfa populations studied (figure 1). Mean genetic diversity ranged from 0.82 in the Kerman population (from subtropical region of Iran) to 0.93 in Ghareyonjeh population from Hamadan (from semi-cold region of Iran). The genetic similarity among individual plants varied from 0.22 to 0.62, indicating a low level of genetic similarity among individuals within populations.

To study population differentiation and to estimate the percentage of intrapopulation and interpopulation genetic variation based on 68 different alleles, analysis of molecular variance (AMOVA) was performed (data not shown). Significant variation was observed among the studied populations ($F_{ST} = 0.024; P = 0.00081$). Although the variation observed among populations was statistically significant, it only accounted for 2.4% of the overall genetic diversity present within these populations; the remaining 97.6% of the genetic variability was explained by the differences among genotypes within populations. Pairwise comparisons of populations based on the F_{ST} values indicated that the Ghareyonjeh population was significantly differentiated ($P < 0.05$) from all other populations except the Lorestan population. The Kerman population from the subtropical region was significantly different ($P < 0.05$) from the Khorasan and Lorestan populations, which originated from the semi-cold region.

The UPGMA clustering algorithm based on the coancestry coefficient matrix was used to group the alfalfa populations. The resulting dendrogram showed a remarkable distinctness of the Ghareyonjeh population from all other populations (figure 2).

Distribution of genetic diversity across different latitudes

The populations studied originated from different latitudes, ranging from 30°N to 36°N. The average genetic diversity in populations from higher latitudes was greater than that of those from lower latitudes (figure 1). A significant positive correlation ($r = 0.88; P < 0.01$) was found between latitude and genetic diversity of the populations. The cluster analysis based on the coancestry coefficient also confirmed that populations located in the same climatic conditions clustered together.

Discussion

SSR polymorphism and genetic diversity in Iranian cultivated alfalfa

To evaluate the informativeness and efficiency of SSR loci in analysis of genetic diversity as well as population differentiation assessments, the total number of alleles, PIC value, and F_{ST} for each SSR locus were estimated. The three SSR loci with PIC values above 0.80, AFct32, AFca11 and AFctt1, showed a total allele number of 14, 12 and 11, respectively. This indicates that markers with large number of alleles are informative for population studies. The high average number of alleles per locus per plant in alfalfa could be due to very high level of heterozygosity and allogamous nature of cultivated alfalfa. In contrast to cultivated alfalfa, in annual medics only one allele per locus per plant was observed using the same SSR loci (Falahati-Anbaran *et al*. 2006). This could be attributed to autogamous and diploid nature of annual medics. A low frequency of plants with either one allele or four different alleles, compared to those with two or three alleles, and normal distribution of the total number of alleles within each plant suggest that breeding programmes were not carried out on Iranian alfalfa populations. Therefore these alfalfa populations could be considered as valuable genetic material for alfalfa improvement programmes.

The level of genetic diversity in Iranian alfalfa populations (on average, 0.89) was as high as or higher than that

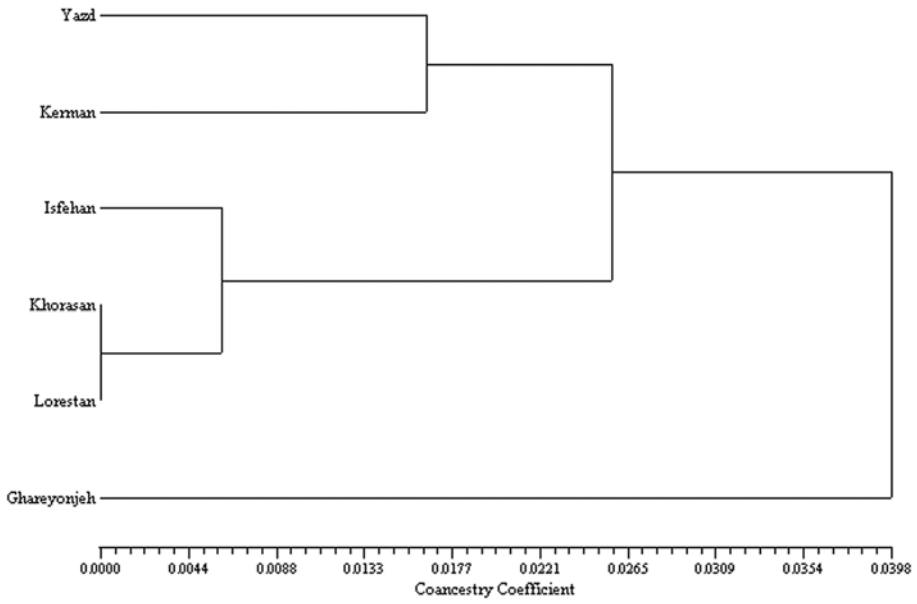

Figure 2. UPGMA clustering of Iranian alfalfa populations using coefficient of coancestry based on 68 SSR alleles.

reported in similar studies on populations from other countries using different marker systems, such as mitochondrial DNA (Muller *et al*. 2001), RAPD (Jenczewski *et al*. 1999; Mengoni *et al*. 2000) and nuclear microsatellite (Mengoni *et al*. 2000) markers. The most likely explanation for the high level of genetic diversity within and among Iranian alfalfa populations is probably that these populations are located in the primary diversity centre of alfalfa. The genetic differences between Italian and Iranian populations may be due to differences in population structure, with the Italian populations being improved cultivars in which breeding processes have decreased the genetic diversity, whereas the Iranian populations are natural populations.

Distribution of genetic diversity across different latitudes

In the absence of other sources of information about the level of genetic diversity in plant germplasm, such as pedigree, and morphological and molecular data, breeders mostly use the geographical distribution of plant materials and the assessment of genetic variation for parental selection. A strong correlation ($r = 0.88, P < 0.01$) was found in the present study between the level of intrapopulation genetic diversity revealed by SSR markers and the geographical distribution of the population. The results indicated a low level of genetic diversity in populations located further from the centres of origin compared to those that are closer to the centres of origin.

Differentiation and relatedness among populations

From SSR analysis, a higher level of genetic variation was observed among individual plants within populations, compared to variation among populations in the Iranian alfalfa populations studied. This suggests that these populations have a high genetic overlap as a result of gene flow through seed and pollen from cultivated populations to others (Muller *et al*. 2001). In this study, although SSR markers could differentiate the populations, overall genetic differentiation among populations was low ($F_{ST} = 0.024, P = 0.00075$) compared to results of other studies.

Determining the relatedness among potential parents is the basis for choosing genetically distant parents in a breeding programme. Cluster analysis indicated the ability and usefulness of SSR markers for studying the differentiation and relatedness among alfalfa populations. The higher distance of the Ghareyonjeh population from the other populations is due to the presence of a specific and exclusive allele at AFca16 locus, with an approximate length of 89 bp. The UPGMA dendrogram clearly shows separation of subtropical and semi-cold populations, such that the Kerman, Yazd, and Isfahan populations, located in adjacent areas and latitudes (30–33°N), cluster together. The alfalfa centre of origin and primary diversity includes the northwestern areas of Iran with semi-cold climatic conditions. The area under Ghareyonjeh is close to the alfalfa centre of origin in Iran and probably other Iranian populations have originated from these regions and were cultivated in other regions in the southern areas of Iran.

In conclusion, the high level of genetic variation within Iranian populations indicates the usefulness of each population, particularly the Ghareyonjeh population, as a valuable genetic source for selection of superior genotypes for development of synthetic varieties with high level of heterosis. The observed specific allele for Ghareyonjeh population could be useful for future studies. This population-specific

marker could be used for identification of the Ghareyonjeh population, which is one of the widely distributed populations in the cold regions, and for determination of the genetic purity of Ghareyonjeh seeds. The level of within-population genetic diversity decreased with distance of the studied populations from the alfalfa centre of origin. The results of the present study indicated greater genetic variation in semi-cold-area populations compared to those from subtropical regions. This study also demonstrated that SSR markers are useful in assessment of genetic relationships among alfalfa populations and population identification, and could, therefore, be used for identifying duplicate accessions, managing conserved germplasm, assessing seed purity in alfalfa collections, and protecting plant breeder rights.

Acknowledgements

The authors thank Profs. E. C. Brummer and Andrea Matzzucato for helpful discussions, guidance and critical reading of the manuscript, Mr Bahram Shariatmadar for assistance with English, and Dr Javad Mozafari for kindly providing seed material.

References

Bassam B. J., Caetano-Anolles G. and Gresshoff P. M. 1991 Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem*. **19**, 680–683.

Botstein D., White R. L., Skolmick M. and Davis R. W. 1980 Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am. J. Hum. Genet*. **32**, 314–332.

Dellaporta S. L., Wood J. and Hicks J. B. 1983 A plant DNA minipreparation: version II. *Plant Mol. Biol. Rep*. **1**, 19–21.

Diwan N., Bhagwat A. A., Bauchan G. B. and Cregan P. B. 1997 Simple sequence repeat DNA markers in alfalfa and perennial and annual *Medicago* species. *Genome* **40**, 887–895.

Diwan N., Bouton J. H., Kochert G. and Cregan P. B. 2000 Mapping simple sequence repeat (SSR) DNA markers in diploid and tetraploid alfalfa. *Theor. Appl. Genet*. **101**, 165–172.

Excoffier L., Smouse P. E. and Quattro M. 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Falahati-Anbaran M., Habashi A. A., Esfahany M., Mohammadi S. A. and Ghareyazie B. 2006 Study of genetic diversity and relationships of diploid and tetraploid annual medics using microsatellite markers. *J. Sci. Tech. Agri. Nat. Res*. **10**, 349–358.

Jenczewski E., Prosperi J. M. and Ronfort J. 1999 Differentiation between natural and cultivated populations of *Medicago sativa* (Leguminosae) from Spain: analysis with random amplified polymorphic DNA (RAPD) markers and comparison with allozymes. *Mol. Ecol*. **86**, 1317–1330.

McCoy T. J. and Bingham E. T. 1988 Cytology and cytogenetics of alfalfa. In *Alfalfa and alfalfa improvement*. Agronomy Monograph 29 (ed. A. A. Hanson, D. K. Barnes and R. R. Hill), pp. 737–776. Madison, USA.

Mengoni A., Gori A. and Bazzicalupo M. 2000 Use of RAPD and microsatellite (SSR) variation to assess genetic relationships among populations of tetraploid alfalfa, *Medicago sativa. Plant Breeding* **119**, 311–317.

Michaud R., Lehman W. F. and Runbaugh M. D. 1988 World distribution and historical development. In *Alfalfa and alfalfa improvement*. Agronomy Monograph 29, (ed. A. A. Hanson, D. K. Barnes and R. R. Hill), pp. 25–92. Madison, USA.

Muller M. H., Prosperi M., Santoni S. and Ronfort J. 2001 How mitochondrial DNA diversity can help to understand the dynamics of wild-cultivated complexes. The case study of *Medicago sativa* in Spain. *Mol. Ecol*. **10**, 2753–2763.

Rohlf F. J. 1998 NTSYS-pc, Numerical taxonomy and multivariate analysis system, version 2.02. Exeter Software, New York, USA.

Schneider S., Kueffer J. M., Roessli D. and Excoffier L. 2000 ARLEQUIN, version 2.000, a software for population genetics data analysis. University of Geneva, Geneva, Switzerland.

Wang Z., Weber J. L., Zhang G. and Tanksley S. D. 1994 Survey of plant short tandem DNA repeats. *Theor. Appl. Genet*. **88**, 1–6.

Received 1 August 2006